

扶桑绵粉蚧转录组分析

胡俊杰^{1,*}, 孟翔², 周佳滨¹, 杨陆兴¹, 刘善海¹, 李润钊¹

(1. 广州大学生命科学学院, 广州 510006; 2. 广东省生物资源与应用研究所, 广州 510260)

摘要:【目的】建立扶桑绵粉蚧 *Phenacoccus solenopsis* Tinsley 的转录组数据库, 揭示扶桑绵粉蚧转录组的整体表达特征。【方法】采用 Illumina HiSeq™ 4000 测序平台开展对扶桑绵粉蚧雌成虫转录组测定, 对原始数据进行过滤和组装。【结果】共获得 58 322 258 条序列读取片段 (reads), 共 9.48 Gb (GenBank 登录号: SAMN06130426) 57 422 032 条有效转录组数据。进一步组装拼接后, 共获得 94 475 个单基因簇 (unigene), 平均长度为 700 bp。将 unigene 与数据库中的序列进行 BLASTX 比对, 成功注释 20 949 个 unigenes, 其中, Nr 注释的 unigenes 与豌豆蚜 *Acyrtosiphon pisum* unigenes 同源性最高, 达 18.69%。扶桑绵粉蚧转录组 unigenes 根据 GO 功能注释大致可分为细胞组分、分子功能和生物过程三大类 55 个分支, 与结合活性、催化活性、细胞进程和代谢进程相关的 unigenes 较多。此外, 本研究还筛选到 20 条与脂类代谢相关的途径和与性信息素代谢相关的序列。并通过与 Nr 和 Swiss-prot 蛋白质数据库比对, 获得了 15 037 条编码序列 (CDS) 片段。【结论】本研究初步阐明扶桑绵粉蚧雌成虫转录组的整体表达模式, 为进一步研究扶桑绵粉蚧的基因功能及性信息素的代谢途径奠定了基础。

关键词: 扶桑绵粉蚧; 转录组; 基因注释; GO 数据库; 编码序列; 性信息素

中图分类号: Q966 **文献标识码:** A **文章编号:** 0454-6296(2017)01-0009-09

Transcriptome analysis of the cotton mealybug, *Phenacoccus solenopsis* (Hemiptera: Pseudococcidae)

HU Jun-Jie^{1,*}, MENG Xiang², ZHOU Jia-Bin¹, YANG Lu-Xing¹, LIU Shan-Hai¹, LI Run-Zhao¹
(1. School of Life Sciences, Guangzhou University, Guangzhou 510006, China; 2. Guangdong Institute of Applied Biological Resources, Guangzhou 510260, China)

Abstract: 【Aim】The objective of this study is to establish the transcriptome database of the cotton mealybug, *Phenacoccus solenopsis* Tinsley, and to reveal the whole expression characteristics of the data. 【Methods】The transcriptome of female adults of *P. solenopsis* was sequenced using an Illumina HiSeq™ 4000 platform, and the raw data were filtered and assembled. 【Results】A total of 58 322 258 reads containing 9.48 Gb (GenBank accession no.: SAMN06130426) valid data (57 422 032 clean reads) were obtained. A total of 94 475 unigenes with a mean length of 700 bp was obtained by further sequence splicing, and 20 949 unigenes were annotated using BLASTX searches against the databases. All unigenes against the Nr database had the highest similarity (18.69%) with those of *Acyrtosiphon pisum*. Moreover, all unigenes were broadly divided into 55 branches of 3 categories (biological processes, cellular components and molecular function categories) using Gene Ontology (GO), and more unigenes were annotated to be related to metabolic process, binding activity, catalytic activity and cellular process. In addition, 20 pathways associated with lipid metabolism pathway and the sequences related to the metabolism of sex pheromone were found. Totally 15 037 coding sequences (CDSs) were obtained using blast in Nr and Swiss-prot protein databases. 【Conclusion】This study preliminarily clarified the whole

基金项目: 广州市属高校科技计划项目 (1201430778); 2016 年度广州大学大学生创新训练项目 (201611078064)

作者简介: 胡俊杰, 男, 1982 年 3 月生, 山西太原人, 博士, 讲师, 研究方向为植物检疫与入侵生物学, E-mail: hujunjie2011@hotmail.com

* 通讯作者 Corresponding author, E-mail: hujunjie2011@hotmail.com

收稿日期 Received: 2016-08-22; 接受日期 Accepted: 2016-12-19

expression pattern of the transcriptome of female adults of *P. solenopsis*, which will provide a foundation for further studying gene functions and the metabolism of sex pheromones in *P. solenopsis*.

Key words: *Phenacoccus solenopsis*; transcriptome; gene annotation; Gene Ontology (GO) database; coding sequence (CDS); sex pheromone

高通量测序技术自 2005 年问世以来 (Margulies *et al.*, 2005), 广泛应用于动植物的基因挖掘、功能鉴定和遗传进化等各个尺度, 已经成为当前生物学的热点研究模式之一 (张棋麟和袁明龙, 2013)。转录组测序 (RNA sequencing) 指利用第二代高通量测序技术进行 cDNA 测序, 是一类专注于功能位点的测序策略, 能全面快速地获取研究材料特定组织在某一状态下的全部转录本信息 (贾新平等, 2014a)。与昆虫全基因组测序相比, 转录组测序具有极高的性价比优势, 而且对目前还没有测定全基因组的昆虫来说, 先测转录组是深入研究昆虫生物学特性分子机理的最佳选择 (张传溪, 2015)。

扶桑绵粉蚧 *Phenacoccus solenopsis* Tinsley 又称棉花粉蚧, 是一种多食性的外来入侵害虫, 目前已经对很多国家的作物和蔬菜等农业生产造成了巨大的破坏 (武三安和张润志, 2009)。自从在美国新墨西哥州第一次发现扶桑绵粉蚧以来 (Tinsley, 1898), 相继在北美洲、南美洲、亚洲以及非洲都有关于该虫发生的报道 (Fuchs *et al.*, 1991; Granara de Willink, 2003; Abbas *et al.*, 2005; Akintola and Ande, 2008)。扶桑绵粉蚧寄主范围广, 能够为害包括棉花在内的 150 余种经济作物和园艺植物 (Arif *et al.*, 2009), 2005 年传入巴基斯坦和印度后对当地的棉花生产造成了极大的损失 (Hodgson *et al.*, 2008; Nagrare *et al.*, 2009)。CLIMEX 模型预测的结果显示该虫是亚洲乃至世界其他棉花产区的重大威胁 (Wang *et al.*, 2010)。

生殖方式是影响外来物种成功入侵的重要因素 (万方浩等, 2011), 而对于扶桑绵粉蚧生殖方式的争议依旧存在。有些学者认为孤雌生殖是该虫的主要生殖方式, 在野外条件下也可能营两性生殖 (Vennila *et al.*, 2010); 而卵泡细胞的凋亡和卵吸收的发生等证据表明扶桑绵粉蚧只能进行两性生殖 (Huang *et al.*, 2013)。该虫究竟以何种生殖方式进行入侵和定殖尚存在争议。昆虫体内的脂类具有储存能量等多种功能, 作为昆虫传递信息的重要物质——信息素的主要成分也是脂类的衍生物, 而且昆虫性信息素的前体主要是类脂体 (王荫长, 2001)。能否通过分析脂类代谢过程发现性信息素

的代谢途径, 并且由此证明扶桑绵粉蚧雌成虫通过分泌雌性信息素吸引雄虫完成交配进行两性生殖呢?

本研究采用 Illumina HiSeq™ 4000 测序平台开展转录组测序工作, 将获得的序列信息进行拼接和组装, 并进行基因功能注释和功能分类, 从而全面探讨扶桑绵粉蚧的转录组信息, 在基因组水平上挖掘扶桑绵粉蚧脂类代谢和性信息素代谢相关的重要功能基因, 为证明扶桑绵粉蚧的生殖方式提供证据, 也为进一步相关基因的克隆与表达奠定基础。

1 材料与方法

1.1 供试昆虫

供试扶桑绵粉蚧 *P. solenopsis* 于 2014 年 5 月采自广州市越秀区扶桑 *Hibiscus rosa-sinensis* 绿化带, 带回实验室转移到棉花植株上, 置于人工气候箱 (上海一恒 MGC-250P) 内饲养。在 $28 \pm 1^\circ\text{C}$ 、相对湿度 $\text{RH } 60\% \pm 5\%$ 、光周期 14L:10D 条件下取连续培养 10 代以上, 收集雌成虫虫体用液氮冷冻, 存于 -80°C 备用。

1.2 总 RNA 的提取与检验

按照 Trizol Reagent 方法提取上述样品总 RNA。采用 1% 琼脂糖凝胶电泳检测总 RNA 的完整性, 经 Nanodrop2000 分光光度计 (IMPLEN, CA, USA) 测定 RNA 的纯度 ($\text{OD}_{260}/\text{OD}_{280}$ 值), 依据 Qubit RNA Assay Kit 使用说明定量测定 RNA 浓度, 并用 Agilent 2100 (Agilent Technologies, CA, USA) 精确检测 RNA 的完整性。

1.3 cDNA 文库构建和测序

取 $1.5 \mu\text{g}$ 检测合格的 RNA 样品进行文库构建: 使用带有 Oligo (dT) 的磁珠富集 mRNA; 加入 fragmentation buffer 将 mRNA 打断成短片段, 以 mRNA 目的片段为模板, 用六碱基随机引物合成 cDNA 第 1 链, 然后加入缓冲液、dNTPs、DNA polymerase I 和 RNase H 合成 cDNA 第 2 链, 再用 AMPure XP beads 纯化双链 cDNA。纯化的双链 cDNA 先进行末端修复、加 A 尾并连接测序接头, 再

用 AMPure XP beads 进行片段大小选择。通过 PCR 扩增和 AMPure XP beads 纯化 PCR 产物得到 cDNA 文库。文库构建完成后, 先使用 Qubit2.0, Agilent 2100 和 Q-PCR 方法对文库进行检测。最后将检验合格的 cDNA 文库在 Illumina HiSeq™ 4000 测序平台上进行测序。cDNA 文库构建与测序由北京诺禾致源生物信息科技有限公司协助完成。

1.4 转录组组装

测序获得的原始测序序列(raw reads)经过处理去除序列接头、poly-N 和低质量测序序列, 获得过滤后的测序数据(clean data)。同时, 在高质量的 clean data 基础上计算 Q20 (Phred 数值大于 20 碱基占总体碱基的百分比)、Q30 (Phred 数值大于 30 碱基占总体碱基的百分比)、GC 含量(碱基 G 和 C 的数量总和占总的碱基数量的百分比)和重复序列水平。获得高质量的测序数据后, 采用 Trinity 软件对 clean reads 进行拼接、过滤和组装。

1.5 基因功能注释和分类

使用 BLAST 程序将拼接得到的 unigene 与下列数据库进行比对, 获得基因的功能注释信息。数据库分别为: NCBI (<https://www.ncbi.nlm.nih.gov/>) 蛋白数据库 (NCBI non-redundant protein sequences, Nr); NCBI 核酸序列数据库 (NCBI non-redundant nucleotide sequences, Nt); 蛋白家族 (protein family, Pfam); 基因本体论 (gene ontology, GO)。根据 NCBI 数据库的功能注释信息, 使用 Blast2 GO 软件得到 unigene 的 GO 条目, 然后用 WEGO 软件对所有的 unigene 进行 GO 功能分类统计 (Conesa *et al.*, 2005; Ye *et al.*, 2006)。

1.6 编码序列 (coding sequence, CDS) 预测与分析

将扶桑绵粉蚧转录组中的 unigenes 与 Nr 蛋白数据库和 Swiss-prot 蛋白数据库 (a manually annotated and reviewed protein sequence database) 进行比对, 若比对上, 则从比对结果中提取转录本的开放阅读框 (open reading frame, ORF) 信息, 并按照标准密码子表将编码区序列翻译成氨基酸序列 (按照 5' - 3' 的顺序); 若未比对上, 则采用 ESTscan 软件预测该 unigene 的 ORF, 得到这部分基因编码的核酸序列和氨基酸序列。

2 结果

2.1 扶桑绵粉蚧雌成虫转录组数据的组装

采用 Illumina HiSeq™ 4000 高通量测序平台对

扶桑绵粉蚧雌成虫转录组进行测序, 共获得 58 322 258 条测序序列 (reads), 通过过滤去杂 (去除接头序列、质量差的序列等) 后, 获得 57 422 032 条 clean reads, 共 9.48 Gb (GenBank 登录号: SAMN06130426) 有效转录组数据。其中, GC% 含量为 38.62%, Q20 为 96.83%, 碱基 Q30 为 92.03%。

通过 Trinity 软件对所获得的 reads 进行组装 (图 1), 共得到 114 648 个转录本 (transcript), 序列信息达到 101 369 057 bp, 平均长度为 884 bp, N50 (覆盖 50% 所有核苷酸的最大序列重叠群长度) 长度为 1 920 bp, 长度为 200 ~ 300 bp 的 transcript 数量最大, 占总体的 36.41%。在 transcript 基础上进一步组装获得 94 475 条 unigenes, 平均长度为 700 bp, N50 长度为 1 297 bp。转录组的长度分布特征分析见图 1, 长度为 200 ~ 300 bp 的 unigene 所占比例最大, 占总体的 41.54%。

2.2 Unigene 的序列注释与相似性分析

利用 BLASTX 程序将获得的 unigenes 序列与 Nr, Nt, Pfam, Swiss-prot 和 GO 数据库进行对比, 得到扶桑绵粉蚧转录组 unigene 的功能信息。在组装获得的 94 475 条 unigenes 中成功注释 20 949 条, 所占比例为 22.17%; GO 数据库注释 15 994 条 unigenes, 占总体的 16.93%, 比例最高; Pfam 数据库注释 15 877 条 unigenes, 占总体的 16.81%; Nr 数据库注释 15 015 条 unigenes, 占 15.89%; Swiss-prot 数据库注释 11 250 条 unigenes, 占总体的 11.90%; Nt 数据库成功注释的 unigenes 不到 10% (3.59%)。

通过与 Nr 库进行比对注释, 扶桑绵粉蚧转录组注释到其他物种的基因序列共 15 015 条。其中与豌豆蚜 *Acyrtosiphon pisum* 的相似基因序列最多, 所占比例为 18.69%; 其次为内华达古白蚁 *Zootermopsis nevadensis*, 所占比例为 14.86%; 其他相似性序列数量大于 2% 的物种有赤拟谷盗 *Tribolium castaneum* (8.74%), 柑橘木虱 *Diaphorina citri* (6.44%), 丽蝇蛹集金小蜂 *Nasonia vitripennis* (3.77%), 体虱 *Pediculus humanus* (2.77%), 家蚕 *Bombyx mori* (2.25%) 和毁侧沟茧蜂 *Microplitis demolitor* (2.09%), 其他物种占 40.39% (图 2)。

2.3 Unigene 的功能分类

GO 数据库分别描述基因参与的生物过程 (biological process)、所处的细胞组分 (cellular component) 及具有的分子功能 (molecular function) 3 个方面。本研究中共有 15 994 条 unigenes 得到了 89 768 个功能注释, 平均每条 unigene 5.61 个 GO

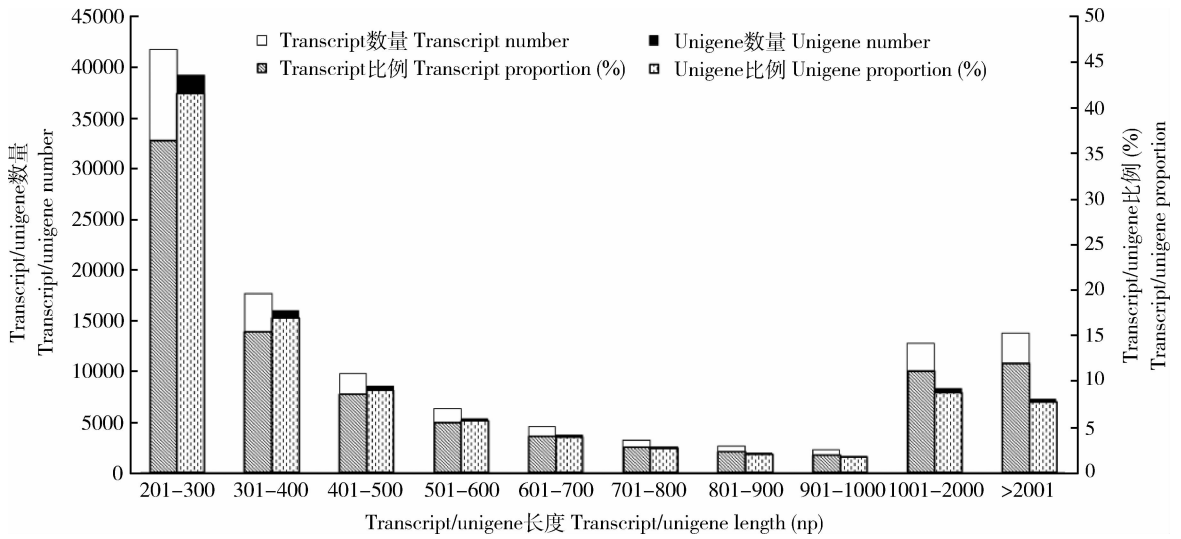


图 1 扶桑绵粉蚧雌成虫转录组 transcript 和 unigenes 数据组装质量统计

Fig. 1 Data assembly for transcript and unigenes in the transcriptome of female adults of *Phenacoccus solenopsis*

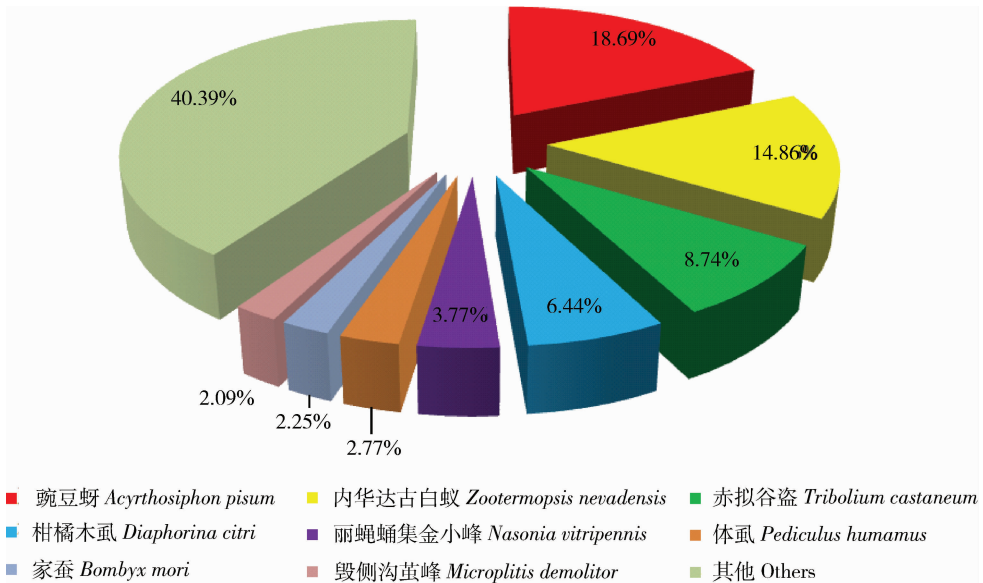


图 2 扶桑绵粉蚧雌成虫与其他物种 Nr 数据库转录组数据比对相似度

Fig. 2 Similarity of transcriptome data of female adults of *Phenacoccus solenopsis* with those of other species in Nr database

注释。其中,19 291 条注释归属于分子功能;27 758 条注释归属于细胞组分;42 719 条注释归属于生物过程。上述 3 大功能可被划分为更为详细的 55 个第 2 层级的描述,分别包含了 14,18 和 23 个第 3 层级的描述(图 3)。结合活性(8 575 个)、催化活性(6 647 个)、单细胞有机体进程(6 895 个)、细胞进程(9 418 个)和代谢进程(8 544 个)相关的 unigenes 较多,金属伴侣蛋白活性(4 个)、调节活性受体(7 个)、核状小体(5 个)、生物相(2 个)和节律进程(9 个)相关的 unigenes 较少。

2.4 脂类代谢相关基因和性信息素相关基因的注释

在扶桑绵粉蚧雌成虫转录组中共筛选到 20 条脂类代谢途径(图 4)。其中,与脂肪酰基还原酶辅酶 A(乙醇形成)活性相关的基因序列最多,共 21 条;与雌性激素代谢过程、雄性激素代谢过程和甾类激素代谢过程相关的序列为 18 条并且是完全相同的基因序列(表 1)。

通过对 18 条基因序列进行 Blast 比对,发现这些序列中有 17 条与脂肪酰基辅酶 A 的代谢相关,

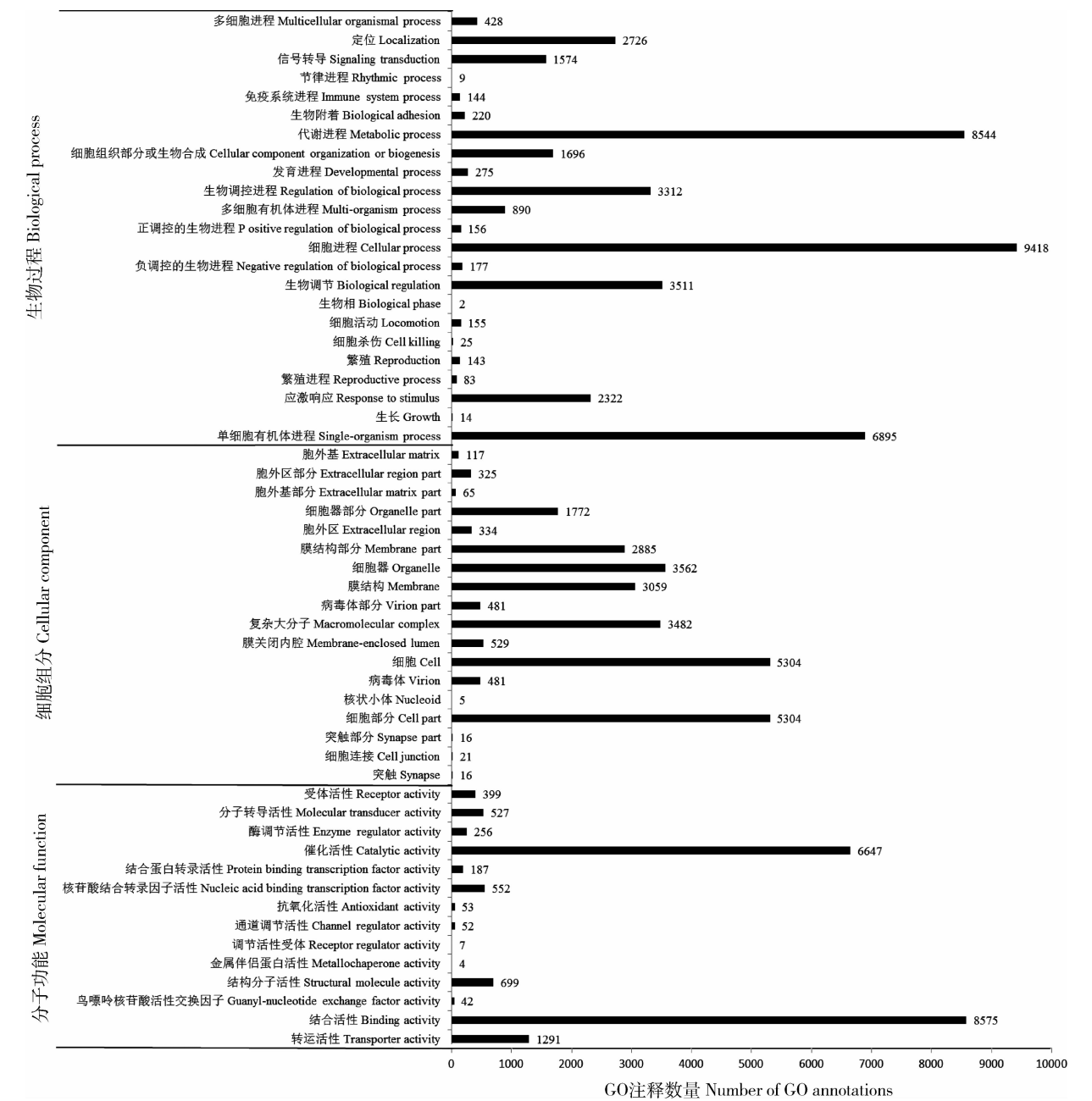


图3 扶桑绵粉蚧雌成虫 unigenes 的 GO 分类

Fig. 3 Gene Ontology categories of female adults of *Phenacoccus solenopsis*

序列 c18425_g1, c33194_g1 和 c26635_g1 与已发表的扶桑绵粉蚧基因 (Li *et al.*, 2016) *PsFAR I* (GenBank ID: ANN23959.1) 一致性较高, 并且 c18425_g1 与 *PsFAR I* 的一致性高达 100%。

2.5 Unigene 的 CDS 预测

按照 Nr 和 Swiss-prot 的优先级顺序, 将扶桑绵粉蚧雌成虫转录组 unigenes 与上述蛋白数据库 blast 比对, 获得 15 037 条 CDS 序列片段, 进一步将比对不上的 unigenes 利用 ESTscan 分析预测获得 6 981 条 CDS 序列片段。在所有 22 018 条 CDS 片段中,

长度在 0 ~ 200 nt 的最多, 有 5 675 条; 其次是 200 ~ 400 nt (5 099 条) 及 400 ~ 600 nt (2 039 条), 其余长度区间的 CDS 都不足 2 000 条 (图 5)。

3 讨论

新一代高通量测序技术具有数据量大、速度快、成本低和效率高的特点, 目前已广泛应用于昆虫分子标记 (Schwarz *et al.*, 2009)、抗药性相关基因的挖掘表达 (Wang *et al.*, 2011) 以及昆虫与其他生物互

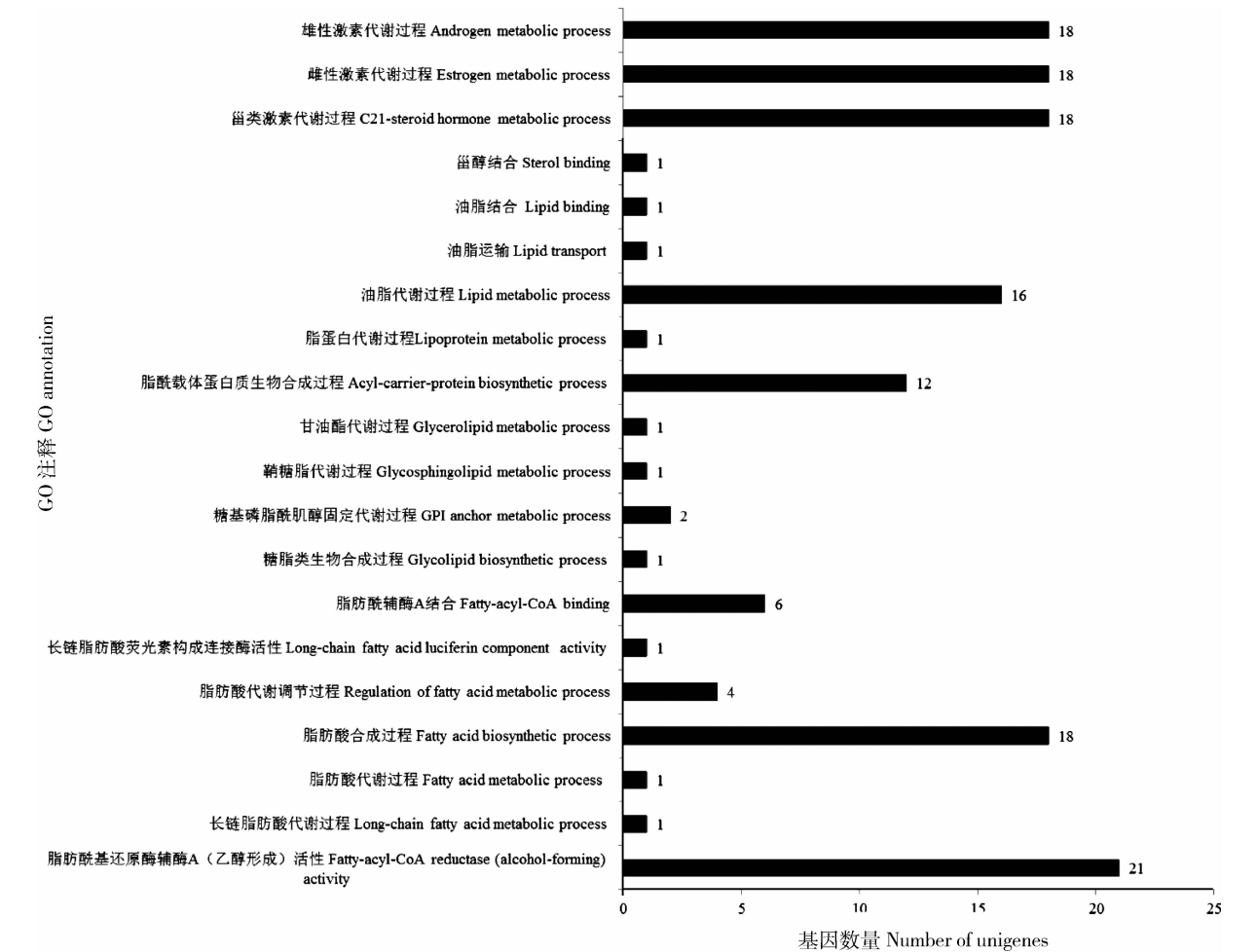


图 4 扶桑绵粉蚧雌成虫脂类代谢相关 unigene 的 GO 注释

Fig. 4 Gene Ontology categories of unigenes related with fatty metabolic process in female adults of *Phenacoccus solenopsis* unigenes

表 1 扶桑绵粉蚧雌成虫性信息素代谢过程相关基因

Table 1 Genes related to sex pheromone metabolic process in female adults of <i>Phenacoccus solenopsis</i>								
Unigene 编号	长度 (bp)	功能推断	物种	NCBI 编号	分数	E 值	一致性 (%)	
Unigene ID	Length	Putative function	Species	NCBI accession no.	Score	E-value	Identity	
c225_g1	2 100	Male sterility domain-containing protein	<i>Pediculus humanus corporis</i>	XP_002426991.1	510	1e-171	49	
c6645_g1	1 577	Fatty acyl-CoA reductase 10	<i>Helicoverpa assulta</i>	AKD01788.1	73.2	2e-12	55	
c18425_g1	2 565	Fatty acyl-CoA reductase 1	<i>Phenacoccus solenopsis</i>	ANN23959.1	1 083	0	100	
c35953_g1	4 281	Fatty acyl-CoA reductase 5	<i>Helicoverpa armigera</i>	AKD01766.1	405	1e-123	45	
c37348_g1	2 065	Fatty acyl-CoA reductase 13	<i>Helicoverpa assulta</i>	AKD01791.1	380	1e-119	43	
c43895_g1	1 911	Fatty acyl-CoA reductase	<i>Ericerus pela</i>	AGK27745.1	570	0	52	
c33194_g1	2 068	Fatty acyl-CoA reductase 1	<i>Phenacoccus solenopsis</i>	ANN23959.1	522	2e-176	55	
c26635_g1	2 490	Fatty acyl-CoA reductase 1	<i>Phenacoccus solenopsis</i>	ANN23959.1	599	0	56	
c38235_g1	2 936	Fatty acyl-CoA reductase	<i>Anopheles darlingi</i>	ETN62440.1	468	6e-152	53	
c35267_g1	2 269	Fatty acyl-CoA reductase 1	<i>Camponotus floridanus</i>	EFN66140.1	632	0	62	
c29157_g1	2 388	Fatty-acyl CoA reductase 5	<i>Danaus plexippus</i>	EHJ72233.1	310	1e-92	35	
c16346_g1	2 509	Fatty acyl-CoA reductase	<i>Ericerus pela</i>	AGK27745.1	528	6e-177	51	
c38946_g3	3 858	Fatty acyl-CoA reductase	<i>Ericerus pela</i>	AGK27745.1	467	2e-148	48	
c49387_g1	2 182	Fatty acyl-CoA reductase 8	<i>Helicoverpa assulta</i>	AKD01786.1	512	4e-172	54	
c18290_g1	1 919	Fatty acyl-CoA reductase 1	<i>Camponotus floridanus</i>	EFN66142.1	443	3e-145	42	
c10213_g1	1 178	Fatty acyl-reductase 1	<i>Lasius niger</i>	KMQ98413.1	132	1e-32	42	
c86525_g1	2 051	Fatty acyl-CoA reductase	<i>Ericerus pela</i>	AGK27745.1	660	0	61	
c67509_g1	2 247	Fatty acyl-CoA reductase	<i>Ericerus pela</i>	AGK27745.1	535	0	50	

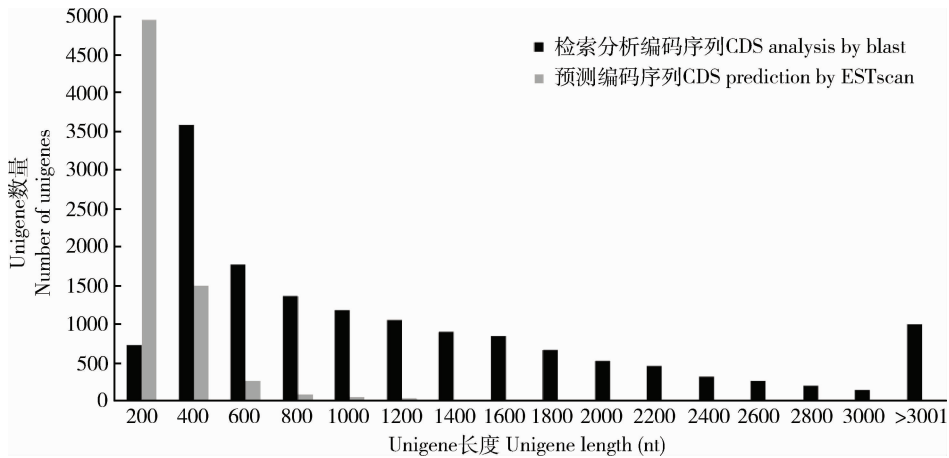


图5 编码序列(CDS)长度分布图

Fig. 5 Coding sequence (CDS) length distribution

作研究等方面 (Zhu *et al.*, 2013)。本研究利用 Illumina 测序技术对扶桑绵粉蚧雌成虫的转录组进行测序和分析,揭示其转录组的整体表达模式。共获得 94 475 条高质量 unigenes,平均长度为 700 bp, N50 值为 1 297 bp。一般认为 N50 值越大就表示得到的长片段越多 (贾新平等, 2014b),且不小于 800 bp 就说明组装得到序列完整性较好;碱基 Q30 值为 92.03%, Q30 在 80% 以上就认为测序质量可靠 (王兴春等, 2015)。上述结果表明本研究测序数据组装的质量和长度满足转录组分析的基本要求,为进一步发掘重要的功能基因奠定了基础。

本研究利用 Nr, Nt, Pfam, Swiss-prot 和 GO 数据库对所获得的 unigene 进行 BLASTX 比对分析,成功注释 20 949 条 unigene,仍有 73 526 条 unigene 无匹配结果,这可能与序片段过短和缺乏注释信息有关 (魏利斌等, 2012)。通过与 Nr 数据库进行同源序列比对,注释到豌豆蚜 *A. pisum* 的序列最多,这是因为豌豆蚜具有丰富的基因组信息 (The International Aphid Genomics Consortium, 2010)并且两种生物的亲缘关系较近。GO 是一个国际化的基因功能分类体系,能够全面描述生物体中基因和基因产物的属性,从宏观上认识该物种的基因功能分布特征,其第 1 层级的描述均分为生物过程、细胞组分及具有的分子功能 3 个方面,而第 2 层级描述在不同物种间还是存在一定差异,如本文中序列归类为 55 个第 2 层级的描述,而在荔枝蒂蛀虫有 47 个第 2 层级的描述 (孟翔等, 2016)。在第 3 层级的描述中,扶桑绵粉蚧的结合活性和催化活性的 GO 注释较多,这与大垫尖翅蝗 *Epacromius coerulipes* (金永玲等, 2015)和甘薯象甲 *Cylas formicarius* (Ma

et al., 2016)的研究结果较为一致。

昆虫性信息素的前体主要来自于寄主植物或者特殊腺体中的类脂体 (闫凤鸣, 2011)。本文通过分析 GO 功能注释,发现扶桑绵粉蚧 20 条脂类代谢途径,其中,雌性激素代谢过程、雄性激素代谢过程和甾类激素代谢过程都注释到相同的 18 条序列。本研究所用材料是雌成虫,这些序列仅仅参与雌性激素代谢过程,合成雌性性信息素吸引雄虫而完成两性生殖吗? 这种推断与在扶桑绵粉蚧广州种群的研究中普遍认为该虫仅能进行两性生殖 (关鑫等, 2011; 胡俊杰, 2012)一致。有文献表明脂肪酰还原酶 (fatty-acyl reductase) 和酰基辅酶 A 去饱和酶 (acyl-CoA desaturase) 在家蚕性信息素合成中起着重要作用 (Moto *et al.*, 2004; Ohnishi *et al.*, 2006)。在扶桑绵粉蚧中, Li 等 (2016) 通过转录组测序发现 *PsFar I* 和 *PsFar II* 的片段,经过克隆获得了 *PsFar I* 和 *PsFar II* 的全长序列,并认为这两个基因具有解毒抗药的功能,但并未描述其参与性信息素合成的功能。基因 *PsFar I* 与本文从性信息素代谢筛选到的序列 c18425_g1, c33194_g1 和 c26635_g1 具有很高的 consistency,发掘该基因的功能对于阐明扶桑绵粉蚧性信息素的代谢途径具有一定的意义。

转录组测序产生的序列所包含的 CDS 的完整性将直接影响到转录组的分析和使用。本研究中还尚未涉及利用转录起始位点 (TIS) 和位点转录终止位点 (TTC) 来评估转录组的拼接质量,这也是今后对扶桑绵粉蚧转录组进行深入分析的工作之一。

转录组是昆虫功能基因研究的重要手段,许多昆虫生物学特性解析与比较都可以从转录组测序并分析差异基因入手。目前,以扶桑绵粉蚧转录组测

序为基础进行的基因功能的研究已经有少量报道,如利用转录组测序找到 1 781 个 SSR 位点,为 481 个 unigenes 成功设计引物,证明利用转录组数据开发 SSR 标记可行(罗梅等, 2014);从转录组数据中获得 12 条化学感受蛋白基因,扶桑绵粉蚧化学感受蛋白基因与其他昆虫的化学感受蛋白相似性都较高(赵洁和陆永跃, 2015)。本研究尚存在不足之处,如雄虫和 1 龄若虫的数量未能达到转录组测序的要求而没有获得该虫态的数据,不能进行不同性别和不同虫态的转录组数据的差异分析,这是下一步深入研究扶桑绵粉蚧要进行的重要步骤。但此次转录组测序仍建立了扶桑绵粉蚧雌成虫转录组数据库,利用功能注释与分类发现了性信息素代谢相关的序列,并和其他学者的研究成果具有高度一致性。本研究将为扶桑绵粉蚧基因的克隆与差异表达以及分子标记开发提供了极大的方便,为半翅目昆虫的分子生物学研究提供了丰富的数据资源。

参考文献 (References)

Abbas G, Arif MJ, Saeed S, 2005. Systematic status of a new species of the genus *Phenacoccus cockerell* (Pseudococcidae), a serious pest of cotton, *Gossypium hirsutum* L. *Pakistan Entomol.*, 27(1): 83–84.

Akintola AJ, Ande AT, 2008. First record of *Phenacoccus solenopsis* Tinsley (Hemiptera: Pseudococcidae) on *Hibiscus rosa-sinensis* in Nigeria. *Agr. J.*, 3(1): 1–3.

Arif MI, Rafiq M, Ghaffar A, 2009. Host plants of cotton mealybug (*Phenacoccus solenopsis*): a new menace to cotton agroecosystem of Punjab, Pakistan. *Int. J. Agric. Biol.*, 11: 163–167.

Conesa A, Götz S, García-Gómez JM, Manuel T, Montserrat R, 2005. Blast2 GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, 21: 3674–3676.

Fuchs TW, Stewart JW, Minzenmayer R, Rose M, 1991. First record of *Phenacoccus solenopsis* Tinsley in cultivated cotton in the United States. *Southwest. Entomol.*, 16(3): 215–221.

Granara de Willink MC, 2003. New records and host plants of *Phenacoccus* from Argentina (Hemiptera: Pseudococcidae). *Rev. Soc. Entomol. Argent.*, 62(3–4): 80–82.

Guan X, Lu YY, Zeng L, 2011. Biology of mealybug *Phenacoccus solenopsis* Tinsley from Guangzhou fed on cotton. *China Cotton*, 38(8): 11–13. [关鑫, 陆永跃, 曾玲, 2011. 棉花上扶桑绵粉蚧广州种群生物学特性观察. *中国棉花*, 38(8): 11–13]

Hodgson C, Abbas G, Arif MJ, Karar SS, 2008. *Phenacoccus solenopsis* Tinsley (Sternorrhyncha: Coccoidea: Pseudococcidae), an invasive mealybug damaging cotton in Pakistan and India, with a discussion on seasonal morphological variation. *Zootaxa*, 1913: 1–35.

Hu JJ, 2012. Study on Ecological Characteristics of *Phenacoccus solenopsis* Tinsley and Control Effects of *Aenasius bambawalei* Hayat against Mealybug. PhD Dissertation, South China Agricultural

University, Guangzhou. [胡俊杰, 2012. 扶桑绵粉蚧生态学特性及班氏跳小蜂对其控制效果的研究. 广州: 华南农业大学博士学位论文]

Huang F, Zhang JM, Zhang PJ, Lu YB, 2013. Reproduction of the solenopsis mealybug, *Phenacoccus solenopsis*: males play an important role. *J. Insect Sci.*, 13(1): 1–12.

Jia XP, Sun XB, Deng YM, Liang LJ, Ye XQ, 2014a. Sequencing and analysis of the transcriptome of *Asplenium nidus*. *Acta Horticulturae Sinica*, 41(11): 2329–2341. [贾新平, 孙晓波, 邓衍明, 梁丽建, 叶晓青, 2014a. 鸟巢蕨转录组高通量测序及分析. *园艺学报*, 41(11): 2329–2341]

Jia XP, Ye XQ, Liang LJ, Deng YM, Sun XB, She JM, 2014b. Transcriptome characteristics of *Paspalum vaginatum* analyzed with Illumina sequencing technology. *Acta Prataculturae Sinica*, 23(6): 242–252. [贾新平, 叶晓青, 梁丽建, 邓衍明, 孙晓波, 余建明, 2014b. 基于高通量测序的海滨雀稗转录组学研究. *草业学报*, 23(6): 242–252]

Jin YL, Cong B, Wang LY, Zhang HY, Dong H, 2015. An analysis of the transcriptome of *Epacromius coerulipes* (Orthoptera: Acrididae). *Acta Entomologica Sinica*, 58(8): 817–825. [金永玲, 丛斌, 王丽艳, 张海燕, 董辉, 2015. 大垫尖翅蝗转录组分析. *昆虫学报*, 58(8): 817–825]

Li XL, Zheng TX, Zheng XW, Han N, Chen XX, Zhang DY, 2016. Molecular characterization of two fatty acyl-CoA reductase genes from *Phenacoccus solenopsis* (Hemiptera: Pseudococcidae). *J. Insect Sci.*, 16(1): 1–7.

Luo M, Zhang H, Bin SY, Lin JT, 2014. High-throughput discovery of SSR genetic markers in the mealybug, *Phenacoccus solenopsis* (Hemiptera: Pseudococcidae), from its transcriptome database. *Acta Entomologica Sinica*, 57(4): 395–400. [罗梅, 张鹤, 宾淑英, 林进添, 2014. 基于转录组数据高通量发掘扶桑绵粉蚧微卫星引物. *昆虫学报*, 57(4): 395–400]

Ma J, Wang RY, Li XH, Gao B, Chen SL, 2016. Transcriptome and gene expression analysis of *Cylas formicarius* (Coleoptera: Brentidae) during different development stages. *J. Insect Sci.*, 16(1): 63.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM, 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 437: 376–380.

Meng X, Hu JJ, Liu H, Ouyang GC, Guo MF, 2016. Analysis of the transcriptome and olfaction-related genes of *Conopomorpha sinensis* Bradley (Lepidoptera: Gracilariidae). *Acta Entomologica Sinica*, 59(8): 823–830. [孟翔, 胡俊杰, 刘慧, 欧阳革成, 郭明昉,

2016. 荔枝蒂蛀虫转录组及嗅觉相关基因分析. *昆虫学报*, 59 (8): 823–830]
- Moto KI, Roelofs WL, Hull JJ, Kurata R, Takahashi S, Yamamoto M, Okano K, Imai K, Ando T, Matsumoto S, 2004. Involvement of a bifunctional fatty-acyl desaturase in the biosynthesis of the silkmoth, *Bombyx mori*, sex pheromone. *Proc. Natl. Acad. Sci. USA*, 101 (23): 8631–8636.
- Nagrare VS, Kranthi S, Biradar VK, Zade NN, Sangode V, Kakde GR, Shukla M, Shivare D, Khadi BM, Kranthi KR, 2009. Widespread infestation of the exotic mealybug species, *Phenacoccus solenopsis* (Tinsley) (Hemiptera: Pseudococcidae), on cotton in India. *Bull. Entomol. Res.*, 99: 537–541.
- Ohnishi A, Hull JJ, Matsumoto S, 2006. Targeted disruption of genes in the *Bombyx mori* sex pheromone biosynthetic pathway. *Proc. Natl. Acad. Sci. USA*, 1039 (12): 4398–4403.
- Schwarz D, Robertson HM, Feder JL, Varala K, Hudson ME, Ragland GJ, Hahn DA, Berlocher SH, 2009. Sympatric ecological speciation meets pyrosequencing: sampling the transcriptome of the apple maggot *Rhagoletis pomonella*. *BMC Genomics*, 10: 633.
- The International Aphid Genomics Consortium, 2010. Genome sequence of the pea aphid *Acyrtosiphon pisum*. *PLoS Biol.*, 8 (2): e1000313.
- Tinsley JB, 1898. An ants'-nest coccid from New Mexico. *Can. Entomol.*, 30: 47–48.
- Vennila S, Deshmukh AJ, Pinjarkar D, Agarwal M, Ramamurthy VV, Joshi S, Kranthi KR, Bambawale OM, 2010. Biology of the mealybug, *Phenacoccus solenopsis* on cotton in the laboratory. *J. Insect Sci.*, 10(1): 113–115.
- Wan FH, Xie BY, Yang GQ, 2011. *Invasion Biology*. Science Press, Beijing. 67–78. [万方浩, 谢丙炎, 杨国庆, 2011. *入侵生物学*. 北京: 科学出版社. 67–78]
- Wang XC, Tan HL, Chen Z, Meng LZ, Wang WB, Fan SC, 2015. Assembly and characterization of the transcriptome and development of SSR markers in *Forsythia suspensa* based on RNA-Seq technology. *Scientia Sinica Vitae*, 45(3): 301–310. [王兴春, 谭河林, 陈钊, 孟令芝, 王文斌, 范圣此, 2015. 基于 RNA-Seq 技术的连翘转录组组装与分析及 SSR 分子标记的开发. *中国科学: 生命科学*, 45(3): 301–310]
- Wang XW, Luan JB, Li JM, Su YL, Xia J, Liu SS, 2011. Transcriptome analysis and comparison reveal divergence between two invasive whitefly cryptic species. *BMC Genomics*, 19: 458.
- Wang YC, 2001. *Insect Biochemistry*. China Agriculture Press, Beijing. 86–92. [王荫长, 2001. *昆虫生物化学*. 北京: 中国农业出版社. 86–92]
- Wang YP, Watson GW, Zhang RZ, 2010. The potential distribution of an invasive mealybug *Phenacoccus solenopsis* and its threat to cotton in Asia. *Agric. Forest Entomol.*, 12: 403–416.
- Wei LB, Miao HM, Zhang HY, 2012. Transcriptomic analysis of sesame development. *Scientia Agricultura Sinica*, 45 (7): 1246–1256. [魏利斌, 苗红梅, 张海洋, 2012. 芝麻发育转录组分析. *中国农业科学*, 45(7): 1246–1256]
- Wu SA, Zhang RZ, 2009. A new invasive pest, *Phenacoccus solenopsis*, threatening seriously to cotton production. *Chinese Bulletin of Entomology*, 46(1): 159–161. [武三安, 张润志, 2009. 威胁棉花生产的外来入侵新害虫——扶桑绵粉蚧. *昆虫知识*, 46 (1): 159–161]
- Yan FM, 2011. *Chemical Ecology*. Science Press, Beijing. 48–64. [闫凤鸣, 2011. *化学生态学*. 北京: 科学出版社. 48–64]
- Ye J, Fang L, Zheng H, Chen J, Zhang Z, Wang J, Li S, Li R, Bolund L, Wang J, 2006. WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res.*, 34: W293–W297.
- Zhang CX, 2015. Current research status and prospects of genomes of insects important to agriculture in China. *Scientia Agricultura Sinica*, 48(17): 3454–3462. [张传溪, 2015. 中国农业昆虫基因组学研究概况与展望. *中国农业科学*, 48(17): 3454–3462]
- Zhang QL, Yuan ML, 2013. Progress in insect transcriptomics based on the next-generation sequencing technique. *Acta Entomologica Sinica*, 56(12): 1489–1508. [张麒麟, 袁明龙, 2013. 基于新一代测序技术的昆虫转录组学研究进展. *昆虫学报*, 56(12): 1489–1508]
- Zhao J, Lu YY, 2015. Identification and evolution analysis on chemosensory protein genes in *Phenacoccus solenopsis* Tinsley (Hemiptera: Pseudococcidae). *Guangdong Agricultural Sciences*, (12): 121–127. [赵洁, 陆永跃, 2015. 棉花粉蚧化学感受蛋白基因的鉴定与进化分析. *广东农业科学*, (12): 121–127]
- Zhu JY, Yang P, Zhang Z, Wu GX, Yang B, 2013. Transcriptomic immune response of *Tenebrio molitor* pupae to parasitization by *Scleroderma guani*. *PLoS ONE*, 8: 173–185.

(责任编辑: 马丽萍)